DOCUMENT RESUME

ED 445 064 TM 031 680

AUTHOR Hill, Richard
TITLE A Success Story from Kentucky.
PUB DATE 2000-06-00
NOTE 12p.; Paper presented at the Annual National Conference of the Council of Chief State School Officers on Large-Scale Assessment (30th, Snowbird, UT, June 25-28, 2000).
AVAILABLE FROM For full text: http://www.nciea.org/publications.html.
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Feedback; *Portfolio Assessment; State Programs; *Student Evaluation; Teacher Attitudes; Teacher Role; *Testing Programs; Training; *Writing Tests
IDENTIFIERS *Kentucky; Kentucky Education Reform Act 1990

ABSTRACT

This paper describes the intentions of educators and policymakers who implemented a new assessment strategy in Kentucky in 1991, along with its associated shortcomings and one notable success. The principle behind the assessment system was that rather than having an on-demand assessment that would inefficiently replicate the information that was already known, Kentucky would have an assessment that simply collected the already-known information. The premise was that teachers were supposed to be running their classrooms in a way that naturally resulted in assessment information, so the need for an external assessment would be changed and reduced if teachers were doing their jobs correctly. In practice, it did not work. Teachers did not change their practices much, and they did not have the needed information. Nor was there much real improvement in student achievement. The system evolved into one in which the statewide assessment program began to tell teachers what to do and how to do it. Only one part of the assessment really worked as anticipated: the writing portfolios. One reason this assessment came to work so well is that feedback from an early set of portfolios that seemed to be "way out of line" was extensively discussed and analyzed. Teacher training was revised, and teachers reduced scoring errors and elicited better work from students. Several reasons why this improvement in assessment and achievement was not possible in other subject areas are discussed. The experience of Kentucky does show that, where there is the right convergence of training, resources, and will, it is possible to create significant gains in learning. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

R. Hill

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 445 064

1

A Success Story from Kentucky

Richard Hill
The National Center for the Improvement of Educational Assessment, Inc.

Paper presented at the 30th Annual National Conference on Large-Scale Assessment,
Council of Chief State School Officers
Snowbird, Utah

June 25-28, 2000

Background

This paper began as a suggestion from Jim Popham. He proposed a session for the CCSSO conference in which three people would provide their idea of what an "instructionally illuminating" assessment might be. I believe his intent was to have us present the types of questions that we felt would provide guidance to teachers—that having viewed these questions, teachers would be able to develop instructional activities that would more effectiye.

My bias is that the charge he gave us was too limiting. It presumes that external assessments—assessments that come from sources outside the classroom—provide significant information to teachers. It also assumes that the more traditional mode of assessment—the administration of a test after instruction to see what students have learned—is a sufficiently efficient approach to instruction.

My bias is that effective assessment occurs *during* learning. The key to increased levels of achievement, in my opinion, lies in feedback—and more specifically, in the volume and precision of targeted feedback that we provide to learners. I conducted a session at CCSSO last year in which I demonstrated that principle in an active workshop for participants. The way we get lots and lots of quality feedback to participants is to have them do something that requires them to *use* the skills we would hope for them to learn, and to have them do it in a manner that makes it possible to observe the degree to which the learner has acquired the necessary skills. The feedback so provided, when combined with a clear understanding of the content and performance standards desired, makes it clear what a learner has mastered, and what the next steps towards mastery will be. That, to me, is what an "instructionally illuminating" assessment would be.

So I am not going to attempt to describe an external assessment that, while perhaps a significant improvement over typical assessments today, is still only a short step from current practice. What I am going to do is describe how we intended to implement the strategy described above when we began the new assessment system in Kentucky in 1991 and relate a success story from that state.

When I started to think about what I wanted to write for this paper, my thoughts immediately went to Ed Reidy. Ed was the Associate, and then Deputy, Commissioner of Education in Kentucky from 1991 until the beginning of 1998. By June of 1998, Ed was working for the Pew Trusts and I called him to talk about my ideas for setting up the Center for Assessment. During that conversation, Ed reminded me of the vision we had shared for assessment in Kentucky and how far short of our

TM031680

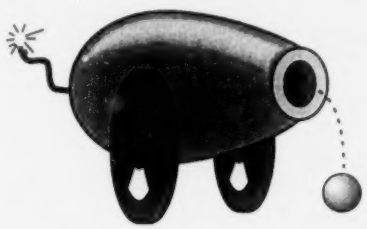
BEST COPY AVAILABLE

ideals we had fallen. The original intent for that assessment program had been that it would be primarily portfolio driven; by 1998, Kentucky's assessment was only a very short distance away from a traditional multiple-choice test with a few open-response questions thrown in. The conversation was a real eye-opener for me; as usual, while the rest of us were wondering how to take the next step forward, Ed was keeping his eyes on the prize. With that reflection, it amazed me how much of that original vision had been lost as we struggled to keep the assessment ship afloat in Kentucky. To commemorate our conversation, I sent Ed the following card—the left side shows the front of the card, and the right side provides what was printed on the inside.

What We Thought We Were Doing

Hey, Claude, let's build us a performance assessment!

What We Actually Did



Sure, Maude. How about adding two short-answer questions to our 60-item multiple-choice test?

Probably because of all the political fallout from the external evaluations that were done in Kentucky, and the subsequent firing of the assessment contractor, there is a common perception that the experience of Kentucky is one that states should avoid, and that nothing done there proved to be a success. That is not an accurate picture. While the overall success of Kentucky's educational reform certainly is arguable[1], at least one effort in Kentucky, the writing portfolios, was clearly successful—that is, it at least partially demonstrated the validity of the approach to instruction outlined above—and one purpose of this paper to make sure that this success story is told at least once. Therefore, I am taking this occasion to (1) reflect on the vision we had, (2) relate the success we had, and (3) discuss the reasons why this success has not been extended.

[1] While it is a digression, it is worthwhile noting that a significant part of the debate in Kentucky has been caused by the failure to distinguish between an assessment system and an accountability system. When we started in Kentucky, we used the terms interchangeably. They are not interchangeable, and the confusion caused by using them that way created serious problems for the assessment system. An assessment system is the process by which we determine students' levels of achievement; an accountability system is the process by which we apply consequences to those levels of achievement and other information we collect about a unit. Our failure to distinguish between those terms led people to believe that changes needed to be made in the assessment program, when, in fact, it generally was operating quite successfully. The majority of changes needed to be made in the accountability system, but because the assessment and accountability systems were not separated, many unnecessary and detrimental changes were made to the assessment system.

The Original Vision

- *The ideal instrument is a test that teachers will want to teach toward.*
- *Teachers should not wait for external forces to tell them how their students are doing. Knowledge of the status of their students is an integral part of their professional responsibility.*
- *Traditional methods of testing have tended to restrain effective education reform, rather than stimulate it.*

The above quotes were all taken from Advanced Systems' proposal for the Kentucky assessment program, written in 1991. What had inspired me to write those words was a presentation delivered by Jack Foster, then the Secretary of Education in Kentucky, at the CRESST conference earlier that year, during which he talked about his vision of a "seamless" assessment. By that, he meant an assessment that would be transparent to students—that the test questions themselves would be so reflective of quality instruction that it wouldn't be clear to students whether they were taking a test or receiving instruction.

Educational reform involves complete rethinking of the roles played by teachers and students. The purpose of the reformed approach is to dramatically improve the efficiency and effectiveness of the educational process. This is achieved, in part, by greatly increasing the amount of information exchanged between student and teacher, and greatly increasing the amount of feedback that students receive. Note that this also greatly increases the amount of feedback the teacher receives, which gives them the information they need to make better decisions about teaching practices.

One of the ways this can happen is by having students be active learners. Active learners produce, as a natural outcome of their activity, artifacts that demonstrate what they can do and what they need to learn next to progress. A primary task of a teacher in a reformed classroom is to use these artifacts skillfully. Observation of these should tell a teacher what skills have been mastered, what areas need to be mastered next for the student to progress, and what teaching approaches have been successful with the student.[2]

The basic principle behind the assessment system we had in mind was that, rather than having an on-demand assessment that would inefficiently replicate the information that was already known, we would have one that simply collected the already-known information. This led to the concept that:

In an ideal system, teachers would tell the state how the school is doing, rather than the state telling the teachers how the school is doing—and they'd be able to do that any day of the year.

---

[2] In fact, the real ideal is for the student to participate in this process, so that he/she is involved in the identification of strengths and weaknesses and can continue to work on strengthening the weak areas. Under the system described, the amount of feedback that the teacher can supply to the student is greatly increased. If the student can self-deliver the feedback, it is increased by yet another factor.

In such an environment, teachers would always know what levels each student had achieved. We wouldn't need to test students to find out if they were Novice, Apprentice, Proficient or Distinguished; we could simply ask the teachers. This approach required, of course, that teachers (a) knew what those levels meant, and (b) were actively engaging their students so they always knew the level of their students' achievements. The premise in Kentucky was that teachers were supposed to be running their classrooms in ways that naturally resulted in assessment information, so the need for an external assessment would be significantly changed and reduced if teachers were doing their jobs correctly.

The Reality

Of course, it didn't work out that way. Teachers didn't change their teaching practices much[3], which meant that they didn't have the needed information available, which meant that the external assessment program had to collect more and more of the information, which meant that it had to become more and more efficient, which meant that all the good forms of assessment, which would have dominated the program if the role of the external assessment was merely to supplement the school's information rather than supplant it, were largely lost because they weren't efficient enough. Also, because teachers didn't change their teaching practices much, there wasn't much real improvement in student achievement to detect, which required the accountability system to be more sensitive to small changes than it was designed to.

At the same time, the fact that we had not clearly distinguished between assessment and accountability exacerbated the situation. When it became clear that the existing accountability system could not detect the small changes in school means that were occurring, it was assumed that the problem was measurement error. In fact it was sampling error[4], but by that time, that wasn't a subtlety that legislators (or even outside evaluators with supposedly the highest credentials of those in the land) were willing to attend to. So more and more multiple-choice questions were added to the mix in an attempt to fix the problem. Including these questions took up more and more of the resources available to implement the assessment program.

As a result, most of initial innovations in Kentucky's assessment system were abandoned. After seven years, the assessment program had deteriorated into a series of external, on-demand tests that were largely multiple-choice with some open-responses questions thrown in. Worse, the idea that assessment should derive first from information the teacher already knew had been lost, and the notion that external tests should be telling teachers what to do and how to do it had gained support. Teachers looked more and more to the statewide assessment program for feedback, which now, because of the increased demand on it, contained far more traditional questions and therefore was less able than ever to provide the needed information.

---

[3] Which is not so say that there wasn't good intent on the part of many teachers statewide. In many cases, they simply didn't know how to change, and often, even why they should change. As will be seen later, a great deal of preparation and training in advance of the desired change is desirable, and probably necessary. This groundwork had not been completed for most teachers in most content areas before the inception of the new accountability system. The significant exception to this was in writing.

[4] A quote from Advanced Systems' original proposal makes it clear that this issue was not an unforeseen problem: *The problem is not the precision of measurement of a particular group; with a range of measures and, if necessary, matrix sampling, one can determine the achievement level of each group with great exactness. The difficulty is the great variability in the groups of students from year to year.* However, because we failed to articulate the issue effectively, it was immaterial that we had anticipated it.

The Success Story—Part 1

However, there was one part of the assessment that worked in the way we envisioned—at least well enough to demonstrate how the whole program could have succeeded given enough time, resources, and true reform in the state's classrooms (which also was a function of time and resources): writing portfolios. Because of the political hullabaloo surrounding Kentucky's accountability system, not many people are aware of the success that writing portfolios enjoyed in Kentucky. But they were a great success story, and a primary purpose for writing this paper is to get that story out.

The story starts during the second year of Kentucky's reform efforts—the 1992-93 school year. Baseline data had been collected during 1991-92. Because the Department and the contractor had limited resources to verify teachers' scores of portfolios, and because there was little reason for teachers to intentionally misscore their students' work, we pretty much accepted the portfolio scores provided by teachers that year as given. But when it was time to assess improvement in 1992-93, we identified a random sample of schools, collected their portfolios, and rescored them. The results were disheartening, to say the least. The rescoring showed that teachers had overstated their students' performance by a wide margin. At all three grades, the scores that teachers had assigned to portfolios were 15-20 points too high—and when the corrected scores were assigned, the average portfolio should have been receiving a score much less than 20. In other words, the errors in teacher-assigned grades were larger than the scores themselves! At that point, we began to question whether it was appropriate to terminate the entire portfolio process.

There was one gleam of hope, however. In addition to collecting portfolios from a random sample of schools, we also had identified another set of schools whose portfolio scores seemed especially out of line, given other information about the school. Our intent was to see whether we could identify the schools doing the worst job of scoring and get the scores of these most serious offenders corrected.

We succeeded in this effort. While the results from the randomly-drawn schools were bad (the results discussed above), the results from the purposively-drawn schools were far worse. Using reverse (or perverse?) logic, we reasoned that since we could identify schools that were doing the worst job at scoring portfolios, there might be some systemic pattern behind scoring errors. If we could find that, perhaps we could salvage the system. With that faint glimmer of hope, we proceeded.

At that point, we did something so dramatic that it almost toppled the system. When it didn't, however, that action proved to be the spark the ultimately led to the success of the writing portfolio system. We sent the results of our rescoring back to the purposively-selected schools and informed them of our intent to change their scores. Teachers and administrators from these schools felt that they had been unfairly singled out. Our data showed that teachers throughout the state had done a poor job in scoring; this group simply had done worse than that, and our sample wasn't large enough to determine whether these schools really were the worst. In addition, because we had no understanding of how the misscoring had occurred, we left open the inference that teachers in these schools had either cheated or were incapable of scoring portfolios. Teachers, quite naturally, took great offense at this. As a result, significant attention was drawn to this issue throughout Kentucky, up to and including the state legislature.

While the immediate effect of the release of the data was quite negative, it ultimately provided significant resources to address the issue and led to a thorough understanding of the causes. There might have been more comfortable ways of tackling this issue, but it is hard to imagine one that could have been more effective.

The release of the "corrected" scores raised howls of protest from these schools. As a result, two people—one from the Department of Education and one from the assessment contractor—traveled throughout the state and held meetings in each of the eight regions shortly after school began in the fall of 1993. At each meeting, they met for an entire day with the teachers from the schools that had been purposively selected.

Each of the eight meetings was virtually identical. For the entire morning, teachers from the schools argued that they had correctly scored the portfolios and that it was the review team that had been in error. As the meeting progressed and their explanations came out, it became clear that the teachers were not using all the categories in the scoring guides to evaluate their students' portfolios. There were six categories that fell into two basic clusters—the primary task of communication, and the "niceties" of writing that one would expect to be polished after several rounds of review and editing. Students' portfolios generally had two characteristics: they were long and they met high standards in the latter category. But when one examined the portfolios in the light of the first group of standards—focus, voice, organization, communication—the writing was generally weak. Teachers had considered only the second cluster of issues in scoring their portfolios, but the scoring review team had correctly included both clusters in their ratings, with far heavier weight on the second. In short, students had produced portfolios that were strong in surface features, but not very good writing.

At the beginning of the meetings, teachers basically were there to tell the state how it had misscored the portfolios during the rescoring process. By about lunch time, the message was starting to get through that despite the teachers *thinking* that they understood the writing standards and scoring guides, they had left out the most important part. By the time the day was over, it was clear to most in attendance that there was much more to the evaluation (and teaching!) of good writing than they had considered in the past, and they now understood what good writing was. More importantly, because most of these teachers had been extensively trained through the Writing Project, they were ready to implement this in their classrooms. This training that they had previously received (Kentucky was an active participant in the Writing Project long before the new state assessment came about) is important for two reasons: first, the fact that the teachers didn't apply the lessons taught by the Writing Project when it came to their students' portfolios highlights the gap between learning and application. Second, even though that instruction wasn't effectively used until after the portfolios had been rescored (and shown to be misscored), it was in place and ready to be used. After participating in just this one-day workshop that provided volumes of feedback to these previously-trained teachers, they were ready to return to their classrooms and make the necessary changes—not just in scoring, but in what they focused on when they were evaluating their students' writing, and therefore, what and how they taught.

Once the dust had settled, it became clear that, despite (or perhaps, because of) volumes of information being distributed to teachers throughout the state, there was little true understanding of the scoring standards. In addition, it appeared that those who had been most involved in the promotion of writing instruction before the implementation of the portfolios were those who paid the least attention to the new materials. One implication of this was that, although much effort had been put into developing a cadre of regional leaders for this process throughout the state, these leaders

were communicating widely discrepant messages back to their regions. Since these leaders had been supposedly delivering "the message of the state," their judgments took on greater weight. When they instructed teachers at the local level about how portfolios should be scored, that information took precedence over anything else the state might be saying. In short, the situation that led to the misscoring of the portfolios resulted from the right information being ignored and the wrong information being closely attended to.

As a result, the state revised several elements related to its training. First, it decided to abandon its multi-tiered "trainer of trainers" model. All control of the portfolio system and the dissemination of all information about how to score portfolios was put in the hands of just the two people who had conducted the regional workshops for the audit schools. When training was done that fall, this pair controlled it all. Since all schools in Kentucky had the capacity to receive interactive video, that system was used extensively. When training was done locally, it was done by videotape, with the local leaders doing nothing but providing materials and turning the videotape machine on and off.

Second, the training materials were extensively revised to include a different mix of sample portfolios. The regional workshops had also led the leaders to understand that poorer teachers were confusing quantity with quality. The portfolios that had received high scores from the teachers but low scores from the audit team typically were quite long, but not focused and written with a purpose and audience. The training materials were revised to emphasize this distinction. We also added hard-to-score portfolios and those near the upper end of each of the performance levels, and training packs so that teachers could test their scoring accuracy themselves. Perhaps the biggest lesson we learned was the truth behind the message we had started with in the beginning, and had not sufficiently applied ourselves—when we had teachers do something, and then saw how they did it, it told us what they knew, what they didn't know, and what we needed to do next to make them more successful at doing what we wanted them to do.

All that occurred in 1993. When it was time to check on the scoring of the portfolios in 1994, we had great trepidation. Because of the political fallout from the original release of the corrected portfolio scores in 1993, the *legislature* established that schools could keep their original scores, if they so desired. As a result, schools had seen that, with sufficient political pressure, they could assign any score whatsoever they desired to their students' work. We braced for the expected onslaught of highly inflated portfolio scores, especially from the schools that had been involved in the audit the previous year.

That isn't what happened at all, however. Table 1 provides the results for the two years of portfolio results for the fourth grade.

Table 1

The Results of Kentucky's First Two Years of Portfolio Scoring—Grade 4

	1992-93		1993-94	
Score	All Schools	Audit Schools	All Schools	Audit Schools
Original	32.7	64.0	37.5	40.4
Corrected	13.3	19.6	28.5	37.4
Difference	19.4	44.4	9.0	3.0

There are several observations that should be made of the data in Table 1. First of all, notice that in 1993, there was a sharp difference in scores between the schools that we purposively selected ("audit schools") and the random sample of all schools in the state ("all schools"). The scoring error ("difference"—the difference between the scores originally submitted by the teachers and those assigned by the rescoring group, or "corrected" scores) was more than twice as much in the audit schools as it was in the general population. This was the fact that we focused on originally. But notice also that, even after correction, the portfolio scores were higher in the audit schools than they were in the general population. The audit schools, as it turned out, tended to be the schools where teachers were more highly trained in the writing process and more committed to their students' writing—which is precisely why the audit results raised such a howl. We had held up to criticism the people who arguably were trying the hardest to do it right.

The 1994 results were a complete surprise to us. Not only hadn't the teachers in the audit schools inflated their scores even more, they turned in results that were, on average, only 3.0 points away from those of the rescoring team. That was a remarkable improvement from the previous year. But note also that the general population cut its scoring error by over half, and also that the corrected scores for everyone (and especially for the audit schools) were considerably higher than they had been the first year.

Just one day of focused feedback had enabled the teachers in the audit schools to draw far better work from their students the second year than they had the first. When equipped with an accurate understanding of how to apply the scoring criteria, teachers were able to do a much better job of helping their students learn and perform. As will be discussed later, it is not true that teachers learned about teaching writing in one day; all that knowledge had been communicated for years through the Writing Project. What happened in that one day was that focused feedback released the information so that teachers could use it effectively for the first time.

The Success Story—Part 2

The great success realized by students and teachers in the audit schools after a limited amount of feedback had an immediate influence on the design of the writing portfolio system from that point forward. Given our renewed understanding of the power of feedback, we wondered whether this was an experience we could duplicate statewide—although we also hoped to avoid the pain associated with the original release of the results for the audit schools.

At the very beginning of the assessment, we had considered the possibility of rescoring portfolios from all schools in the state, and rejected that as too expensive. Given the success with the audit schools, however, that position was reconsidered. We finally settled on a plan that would have us rescore a sample of 25 portfolios from half the schools in the state over the summer of 1994, and the other half in the summer of 1995. Teachers from across the state were recruited for this effort. They were extensively trained. The teachers' scoring accuracy was checked by table leaders, whose accuracy was, in turn, checked by the project managers. Details about these issues are provided in the papers written by Amy Awbrey. Suffice it to say that after the uproar created by the release of the first audit results, great care was taken to ensure that all results produced by these teams could be fully justified.

In this process, a small team of 5-6 teachers reviewed the portfolios from one school at the same time and then immediately discussed their observations. It quickly became clear that, in the

process of reading the portfolios, the team could see what the teachers were trying to accomplish with their students; they therefore had the opportunity to offer suggestions for improving not just the scoring of portfolios, but also the process by which the students were learning to write. As a result, the feedback provided to schools included information not only about the accuracy of the scores they provided, but also about how to more effectively instruct their students in writing. This part of the program was very well received.

By the end of the summer of 1995, therefore, most of the schools in the state had received feedback about how accurately they were scoring their portfolios, and had also received suggestions about how they might improve their instruction in writing. During the summer of 1996, a final rescoring session was conducted. A sample of portfolios was chosen from schools all over the state in order to estimate the amount of scoring error. The results for all years from 1993 to 1996 are shown in Table 2.

Table 2

Four Years' Results for Portfolios—All Grades

Grade	Year	Reported Average	Scoring Error	Real Average
4	1993	31.0	19.4	11.6
	1994	37.6	9.0	28.6
	1995	40.3	10.4	29.9
	1996	39.3	5.1	34.2
8	1993	28.0	15.8	12.2
	1994	31.4	14.1	17.3
	1995	31.8	17.4	14.4
	1996	27.5	4.3	23.2
12	1993	41.1	19.9	21.2
	1994	39.6	17.3	22.3
	1995	38.6	5.4	33.2
	1996	38.9	5.1	33.8

The first column is the data the public saw. These were the results for writing portfolios for the state if one simply took the scores reported by teachers as fact. The second column, however, shows the amount of scoring error. These figures are taken from the re-scoring analyses that were done each summer on portfolios from random sample of schools statewide. The final column, then, simply subtracts the amount of observed scoring error from the reported average to estimate what the average would have been if all portfolios had been accurately scored. The results show statewide improvement over the four years that could be characterized as anywhere from strong to dramatic. Remember that the results in 1993 were, in actuality, baseline data. No one had received any feedback about the accuracy of their scoring before then. Although schools started developing portfolios and reporting those scores in 1992, the feedback system only began after the 1993 scores were received. Half the schools received feedback in 1994, and the other half in 1995. This would explain why, in two of the grades, there was no real decline in scoring error between those two years—the data for each of the years reflects the scores from schools that had not received any feedback. It was not until 1996 that we rechecked the scoring accuracy of teachers who had received feedback. The improvement in scoring accuracy was quite dramatic. As had been true for the audit

BEST COPY AVAILABLE

schools between 1993 and 1994, scoring accuracy increased greatly, and almost immediately, with a minimal amount of feedback. By the end of 1996, the average scoring error was less than 5 points; just three years earlier, it had been almost 20 points.

When the changes in scoring error are taken into account, the improvement in writing performance from 1993 to 1996 is dramatic. At grade 4, scored almost tripled in three years; at grade 8, they almost doubled; and at grade 12, they increased by over 50 percent.

Why Progress Has Been Limited to Writing Portfolios

Why, one might ask, if this was so successful, isn't this success being replicated in other states and in areas other than writing portfolios? Certainly, there are characteristics that are unique to writing and writing portfolios that limited success to this area. Reviewing those will help explain the success with writing, and the failure, to date, to replicate that success with other content areas.

First of all, it is more natural in writing to implement the kinds of teaching practices advocated earlier in this paper. While it may be difficult to think of many situations where students can actively exhibit their learning in, say, social studies, that demonstration is intrinsic to writing—just have them write. Also, in contrast to other area, activity in writing just requires paper and pencil—tools that are already in use in both good and weak classrooms.

Second, it is takes less training and change to efficiently transmit concepts than factual information through this applied approach. While there is a body of factual knowledge to learn about writing, it is considerably smaller than it is for most other content areas. It also is easier to see whether students are acquiring that knowledge (of spelling, grammar, etc.) through writing than it is in other content areas.

Third, there is a high degree of agreement about what constitutes quality writing—and this agreement is shared between the rank and file as well as the leadership in this area. With agreement on the content standards, it is easier to apply more uniform standards of scoring than it is when there is ongoing debate between what is quality work.

Those are all reasons why success in Kentucky was limited just to writing. Efforts to implement portfolios in mathematics never had the same impact, and plans to develop other portfolios were shelved. The reasons why Kentucky's success was not translated to other states include the political fallout from the evaluations of Kentucky's assessment program and failure to get out the message about the parts of the program that had been successful. But there are reasons in addition to these.

First, it is important to note the quantity and quality of in-service that had been on-going in Kentucky for years. Had that training not been in place, it would have been necessary to provide it before teachers would be ready to make the kinds of changes required for this approach.

Second, Kentucky's accountability system provided the need to assure that scores were accurate, and therefore, led us somewhat accidentally into a more complete understanding of the power of feedback. We created the feedback system initially to improve the accountability system; we found that the feedback made us better teachers and made the teachers better learners. If the portfolios hadn't been tied to accountability, I don't think we would have realized this.

Also, the demands of accountability gave us access to resources that otherwise would not have been available. Rescoring a sample of papers from every school was an extraordinarily expensive effort. We believe it proved its worth. The increase in writing effectiveness statewide made the effort worthwhile even if the results hadn't been used for accountability. But it would be difficult to convince a state to invest that much in such an effort if accountability were not a driving factor. In addition, accountability gave a sense of urgency to the quality of the teachers' scoring. Teachers paid attention to the feedback we gave because they knew the results would be used in the accountability system.

Finally, it is worth noting that not even Kentucky has taken full advantage of this learning experience. The gains shown in Table 2 make it clear that not all schools (and perhaps not even most schools) made the necessary changes. An average gain of 20 points does not come about by having all schools increase their performance by 20 points; it comes by having some schools change by far more than that while others have minimal change. Thus, while some teachers benefited greatly from the portfolio experience, many continued to teach in their accustomed fashion, and found the portfolios to be an excessive burden on their "instructional" time.

By 1996, there was enough opposition to the assessment program to cause the Department of Education to change it. Significant changes were made in the entire system, including the writing portfolios. Failure to articulate clearly the success portfolios had been, combined with the confusion between assessment and accountability, led to substantial changes, many of which negated earlier gains.

In summary, it is not easy to make this kind of change, and many forces can serve to stop it. But the experience of Kentucky shows that when there is the right convergence of training, resources and will, it is possible to create significant gains in learning, and appropriate assessment practices play an important role in making that happen.

Don't let it be forgot
That once there was a spot
For one brief shining moment
That was known as Camelot.

References

1. *Advanced Systems' Proposal to Implement the Kentucky Instructional Results Information System*, Advanced Systems in Measurement and Evaluation, Inc., Dover, NH, May, 1991.

2. *1994-95 Writing Portfolio Scoring Analysis*, Amy Awbrey, Advanced Systems in Measurement and Evaluation, Inc., Louisville, KY, October, 1995.

3. *1996 Writing Portfolio Audit: Final Report*, Amy Awbrey, Advanced Systems in Measurement and Evaluation, Inc., Louisville, KY, September, 1996.